3 Basic Simulations

In the basic simulations, we investigate the numerical performance of our multiple testing approach on different fabricated dependence structures where we can control the ground truth parameters. We first simulate \( \theta \) from \( P(\theta; \phi) \) and then simulate \( x \) from \( P(x; \theta; \psi) \) under a variety of settings of \( \theta = (\phi, \psi) \). Because we have the ground truth parameters, we have two versions of our multiple testing approach, namely the oracle procedure (OR) and the data-driven procedure (LIS). The oracle procedure knows the true parameters \( \theta \) in the graphical models, whereas the data-driven procedure does not and has to estimate \( \theta \). The baseline procedures include the BH procedure (Benjamini & Hochberg, 1995) and the adaptive \( p \)-value procedure (AP) (Benjamini & Hochberg, 2000; Genovese & Wasserman, 2004) which are compared by Sun & Cai (2009). We include another baseline procedure, the local false discovery rate procedure (localFDR) (Efron et al., 2001). The adaptive \( p \)-value procedure requires a consistent estimate of the proportion of the true null hypotheses. The localFDR procedure requires a consistent estimate of the proportion of the true null hypotheses and the knowledge of the distribution of the test statistics under the null and alternative hypothesis. In our simulations, we endowed AP and localFDR with the ground truth values of these in order to let these baseline procedures achieve their best performance.

In the simulations, we assume that the observed \( x_i \) under the null hypothesis (namely \( \theta_i = 0 \)) is standard-normally distributed and that \( x_i \) under the alternative hypothesis (namely \( \theta_i = 1 \)) is normally distributed with mean \( \mu \) and standard deviation 1.0. We choose the setup and parameters to be consistent with the work of Sun & Cai (2009) when possible. In total, we consider three MRF models, namely a chain-structured MRF, tree-structured MRF and grid-structured MRF. For chain-MRF, we choose the number of hypotheses \( m = 3,000 \). For tree-MRF, we choose perfect binary trees of height 12 which yields a total number of 8,191 hypotheses. For grid-MRF, we choose the number of rows and the number of columns to be 100 which yields a total number of 10,000 hypotheses. In all the experiments, we choose the number of replications \( N = 500 \) which is also the same as the work of Sun & Cai (2009). In total, we have three sets of simulations with different goals as follows.

Basic simulation 1: We stay consistent with Sun & Cai (2009) in the simulations except that we use the three MRF models. In all three structures, \( (\theta_i) \) is generated from the MRFs whose potentials on the edges are \( \begin{pmatrix} \phi & 1 - \phi \\ 1 - \phi & \phi \end{pmatrix} \). Therefore, \( \phi \) only contains parameter \( \phi \), and \( \psi \) only contains parameter \( \mu \).

Basic simulation 2: One assumption in basic simulation 1 is that the parameters \( \phi \) and \( \mu \) are homogeneous in the sense that they stay the same for \( i(i = 1, ..., m) \). This assumption is carried down from the work of Sun & Cai (2009). However in many real-world applications, the transition parameters can be different across the multiple hypotheses. Similarly, the test statistics for the non-null hypotheses, although normally distributed and standardized, could have different \( \mu \) values. Therefore, we investigate the situation where the parameters can vary in different hypotheses. The simulations are carried out for all three different dependence structures aforementioned. In the first set of simulations, instead of fixing \( \phi \), we choose \( \phi \)'s uniformly distributed on the interval \( (0.8 - \Delta(\phi))/2, 0.8 + \Delta(\phi))/2 \). In the second set of simulations, instead of fixing \( \mu \), we choose \( \mu \)'s uniformly distributed on the interval \( (2.0 - \Delta(\mu))/2, 2.0 + \Delta(\mu))/2 \). The oracle procedure knows the true parameters. The data-driven procedure does not know the parameters, and assumes the parameters are homogeneous.

Basic simulation 3: Another implicit assumption in basic simulation 1 is that each individual test in the multiple testing problem is exact. Many widely used hypothesis tests, such as Pearson’s \( \chi^2 \) test and the likelihood ratio test, are asymptotic in the sense that we only know the limiting distribution of the test statistics for large samples. As an example, we simulate the two-proportion \( z \)-test in this section and show how the sample size affects the performance of the procedures when the individual test is asymptotic. Suppose that we have \( n \) samples (half of them are positive samples and half of them are negative samples). For each sample, we have \( m \) Bernoulli distributed attributes. A fraction of the attributes are relevant. If the attribute \( A \) is relevant, then the probability of “heads” in the positive samples \( p_{+A}^+ \) is different from that in the negative samples \( p_{-A}^- \). \( p_{+A}^+ \) and \( p_{-A}^- \) are the same if \( A \) is non-relevant. For each individual test, the null hypothesis is that the attribute is not relevant, and the alternative hypothesis is otherwise. The two-proportion \( z \)-test can be used to test whether \( p_{+A}^+ - p_{-A}^- \) is zero, which yields an asymptotic \( N(0, 1) \) under the null and \( N(\mu, 1) \) under the alternative (\( \mu \) is nonzero). In the simulations, we fix \( \mu \), but vary the sample size \( n \), and apply the aforementioned tree-MRF structure \( (m = 8, 191) \). The oracle procedure and localFDR only know the limiting distribution of the test statistics and assume the test statistics exactly follow the limiting distributions even when the sample size is small.

Figure 2 shows the numerical results in basic simulation 1. Figures (1a)-(1f) are for the chain structure. Figures (2a)-(2f) are for tree structure. Figures (3a)-(3f) are for the grid structure. In Figures (1a)-(1c),